

ディープラーニングを利用した顔画像による着席状態の集中度解析

1

平石研究室
Shen Jinjian (シン キンケン)
G21606

背景

- 日本国内2020年1月以降,感染者が確認され始め,2月から全国で患者が散発し,政府は全国学校臨時休業を要請した.2022年3月の時点で,世界の感染者の合計数は5億を突破した.
- ユニセフが発表した新しいデータによると,新型コロナウイルスのパンデミックにより,世界で1億7000万人以上の子どもたちの通う学校が約1年間にわたって休校状態となっていた.さらに,約2億2000万人,約7人に1人の子どもが,対面学習の4分の3以上を受けられなかった.

遠隔授業の形式

コロナの影響で対面授業からオンライン授業の割合が多くなった。遠隔授業の形式は、オンライン授業（同期型）とオンデマンド授業（非同期型）の2つに分けることができる。

- ▶ オンライン授業では、ZoomやSkypeのようなテレビ会議システムを利用した遠隔授業である。
- ▶ オンデマンド授業は動画配信サービスを利用し、決められた配信期間中に映像を閲覧する形式の授業である。

遠隔授業の問題

- 基本的には1人で授業を受けるため、怠(なま)けていても誰からも注意されることもない。(オンライン授業)
- 教室のような緊張感がないため、モチベーションを保つのが難しいということもある。(オンライン授業)
- 自己管理ができないとやらないままになってしまう可能性があり。(オンデマンド授業)

目的

5

- ▶ オンライン授業中の学生の状態の把握は難しいのが現状である.
- ▶ 集中や散漫等の状態を確認できれば、授業内容で補足説明が必要な場合や、学生の興味の状態などを把握することが可能となり、授業を効率的に実施（じっし）することができるものと考えられる.

関連研究

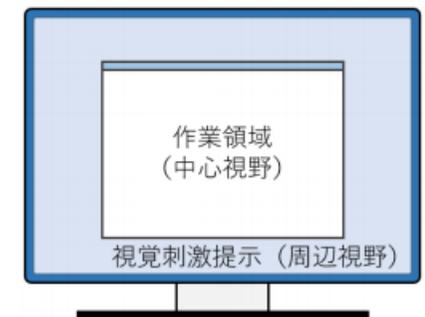
- 集中や散漫等の状態を検出する手段は様々である。例えば、脳波による解析を行なった研究や、座圧センサを利用した着席者の疲労状態の解析や、一点注視型タスクにおける周辺視野（しや）への視覚刺激（しげき）提示が集中度を利用した研究などが存在する。

関連研究：機械学習による座圧センサを利用した着席者の疲労状態の解析（LI 2021）

- 関連研究では座圧センサーを利用して、2人の被験者の面圧中心位置の変化のデータを収集した。
- そのデータの変動の大きさから、集中状態と疲労状態を分類し、3種類の機械学習手法を利用して、授業を受けている時の状況、レポート作成時の状況、コンピュータゲームを実行している時の状況の3つの状態でデータを収集し、それらのデータに対してSVM, RF, また、LSTMによる分析を実施した。
- それにより、3種類の機械学習手法を利用して、それらの状態を自動的に分類することができることを明らかにした。

関連研究：一点注視型タスクにおける周辺視野への視覚刺激提示が集中度に及ぼす影響（桑原 樹蘭 2018）

- ➡ 関連研究では人間の視野には中心視野と周辺視野の2つの領域（りょういき）が存在する，中心視野は解像度（かいぞうど）が高く細部まで認識することを得意としており，周辺視野は対象物の全体像を瞬時に知覚することを得意とする視野範囲である。
- ➡ 関連研究ではタスクにおける視線移動があったために，周辺視野へ視覚刺激を提示し，その刺激量を徐々に減衰（げんすい）させることで集中が増加するのではないかという仮定をもとに実験を行ってきた。
- ➡ 結果：PC作業時の周辺視野に単純な視覚刺激を提示することで作業効率（集中度）を向上するのに効果がある。
- ➡ （中心を見ると、集中状態を判断する；周辺を見ると、散漫状態を判断する。）

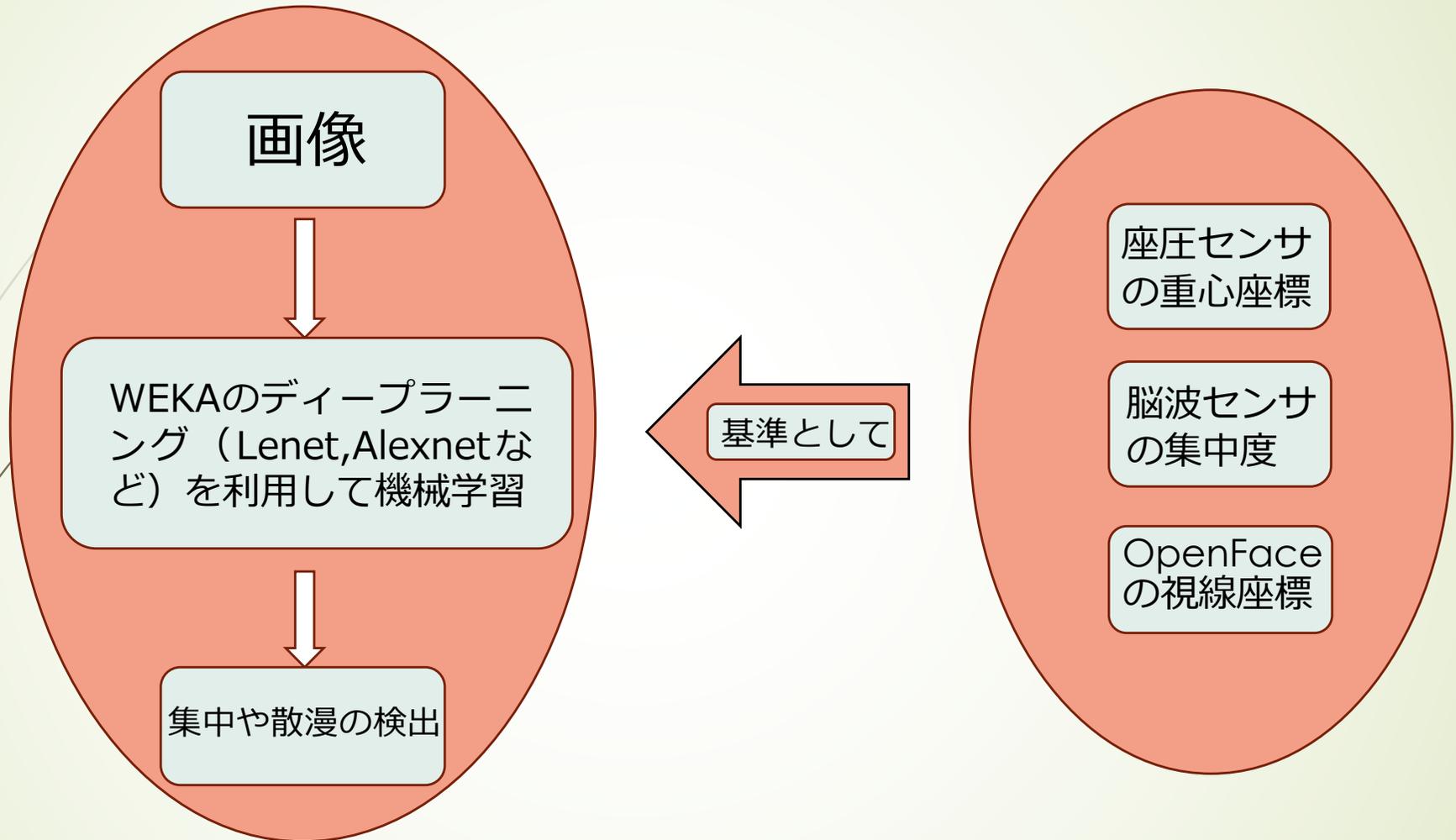


本研究について

- 利用者の集中度の基準となるデータとして、簡易脳波センサと座圧センサを利用した。
- 簡易脳波センサによって利用者の集中や散漫の状態を検出することが可能であり、座圧センサによって利用者の疲労の状態を検出することができる。
- 座圧センサや脳波センサを使わないで、受講者の集中や疲労の状態を把握することが本研究の目的。
 - 着席者の顔画像を収集し、各センサによって得られた集中度との関連性の学習を行なった。
 - 利用者の顔画像を元にディープラーニングを利用した着席者の集中度の解析を試みる。
 - 利用者の視線と顔画像をディープラーニングを分析して、比較する。

使用機材とソフトウェア

- 脳波センサ（ MindWave Mobile ）
- 座圧センサ（ SRソフトビジョン ）
- OpenFace
- WEKA
- 利用したディープラーニング（ LeNet、 AlexNet ）
- RANDOMFORST（ 視線解析利用 ）



脳波センサ

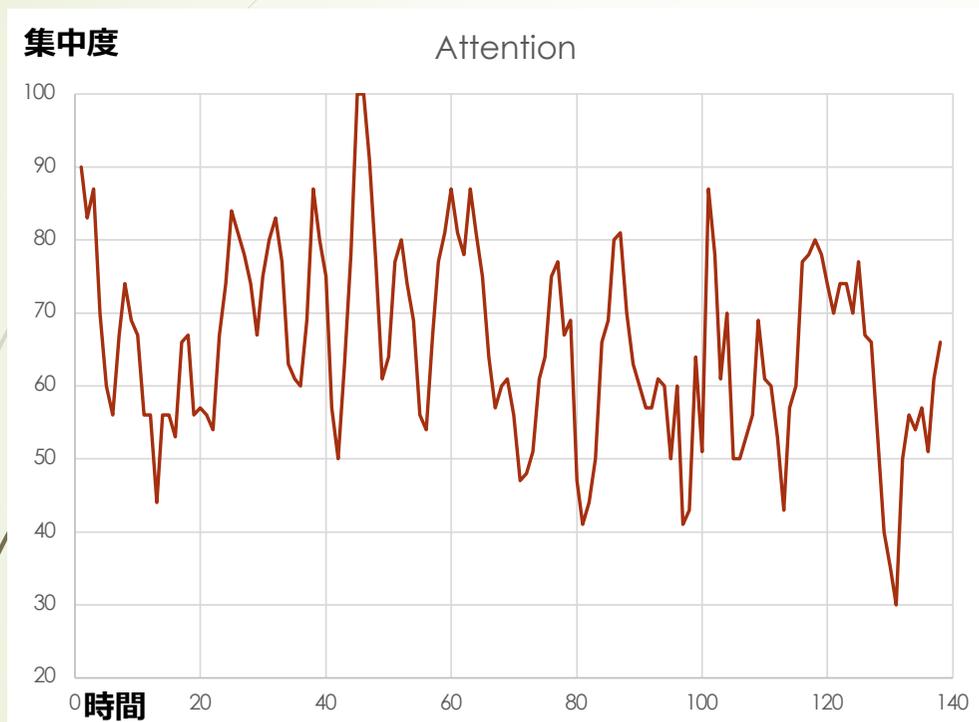
12

本研究では脳波センサとして、MindWave Mobileを利用した。MindWave Mobile はNeurosky社によって開発された脳波センサです。

- ▶ **Attention** 集中度のレベルを測定する。ベータ波を中心とした脳波の変化量からリアルタイムで測定します。
- ▶ **Meditation** リラックス度のレベルを測定します。主にアルファ波を中心とした脳波の変化量からリアルタイム測定します。

Mind Wave Mobile	データ	パワースペクトル
	attention	0-100
	meditation	0-100
	デルタ波	0.5-2.75Hz
	シータ波	3.5-6.75Hz
	低アルファ波	7.5-9.25Hz
	高アルファ波	10-11.75Hz
	低ベータ波	13-16.75Hz
	高ベータ波	18-29.75Hz
	低ガンマ波	31-39.75Hz
	中ガンマ波	41-49.75Hz

脳波データ（集中度）



図は被験者のドラマを視聴（しちょう）する際の集中度を測定したものである。横軸（よこじく）は時間を分単位で示したものであり、縦軸（たてじく）は集中度の平均値を示している。データは1秒間隔で測定しており、毎秒を示している。

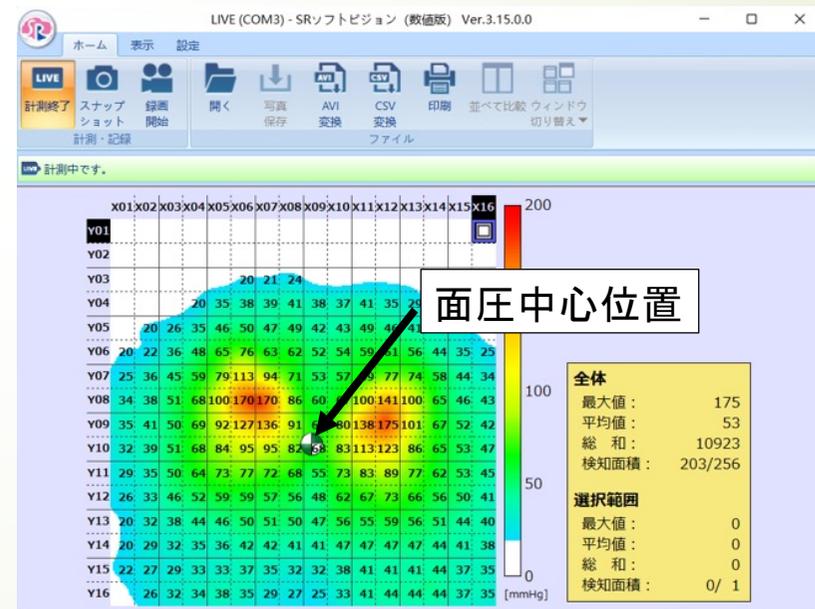
座圧センサ

14

座圧センサとしてSRソフトビジョンを利用した。この座圧センサでは、専用のソフトウェアを利用して圧力分布に加え、圧力値、面圧中心位置をリアルタイムに可視化することができる。データを収集する頻度は自由に設定することが可能である。

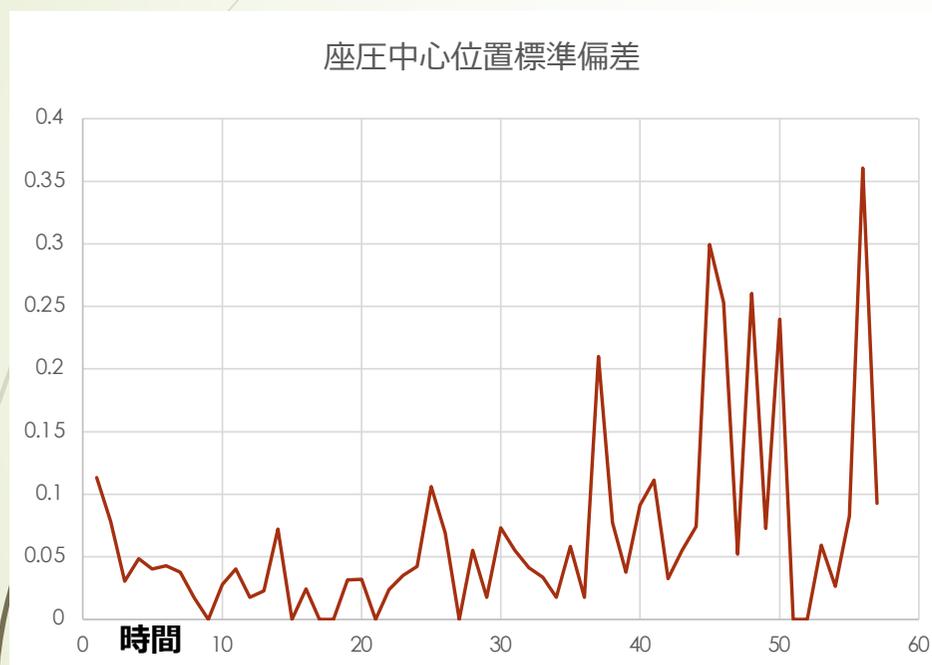


SRソフトビジョン



専用のソフトウェア

座圧中心位置標準偏差による疲労分析

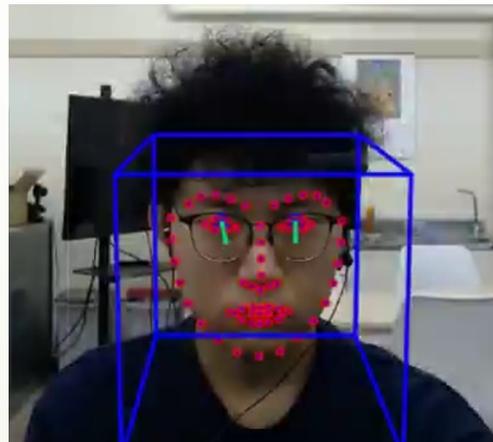


図は被験者がゲームをしている際の座圧を測定したものである。横軸（よこじく）は時間を分単位で示したものであり、縦軸（たてじく）は座圧中心位置標準偏差の平均値を示している。データは5秒間隔で測定しており、1分間の平均値を示している。値は座圧測定画面で示した座標値の平均である。

顔画像と視線

16

本研究では、顔認識と着席者の視線計測のために、Googleによって開発されたFaceNetアルゴリズムを基（もと）にしたOpenFaceを利用した。OpenFaceでは顔の立体的情報を取得し、眼、鼻、あごの輪郭、および、視線データを毎秒30枚顔画像収集することができる。図はOpenFaceの実行の様子である。



WEKA

17

Wekaは、ニュージーランドのワイカト大学で開発した機械学習ソフトウェアで、Javaで書かれている。データ解析と予測モデリングのための視覚化ツールとアルゴリズムの集合体である。

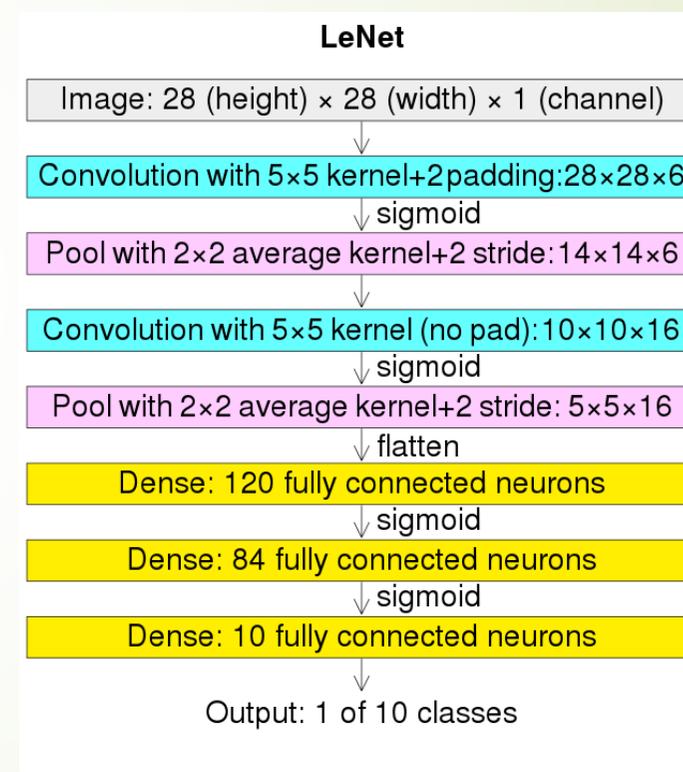


WEKAのソフトウェアの画面

本研究利用したディープラーニング

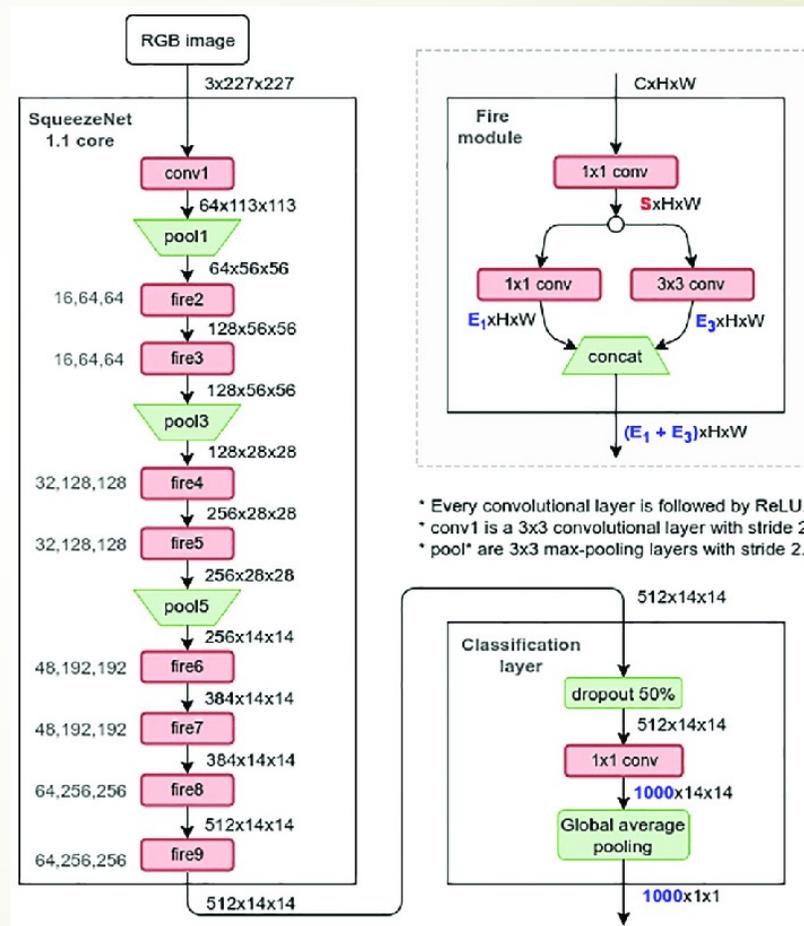
▶ LeNet

LeNet は畳み込みニューラルネットワークの最初の提案および実装であり、初めてのディープラーニング化を行った研究であった。LeNet の構造は全体で畳み込み層が2層、サブサンプリング層が2層、全結合層が3層、そして出力層という構造であるため、7層から成り立っている。



本研究利用したディープラーニング

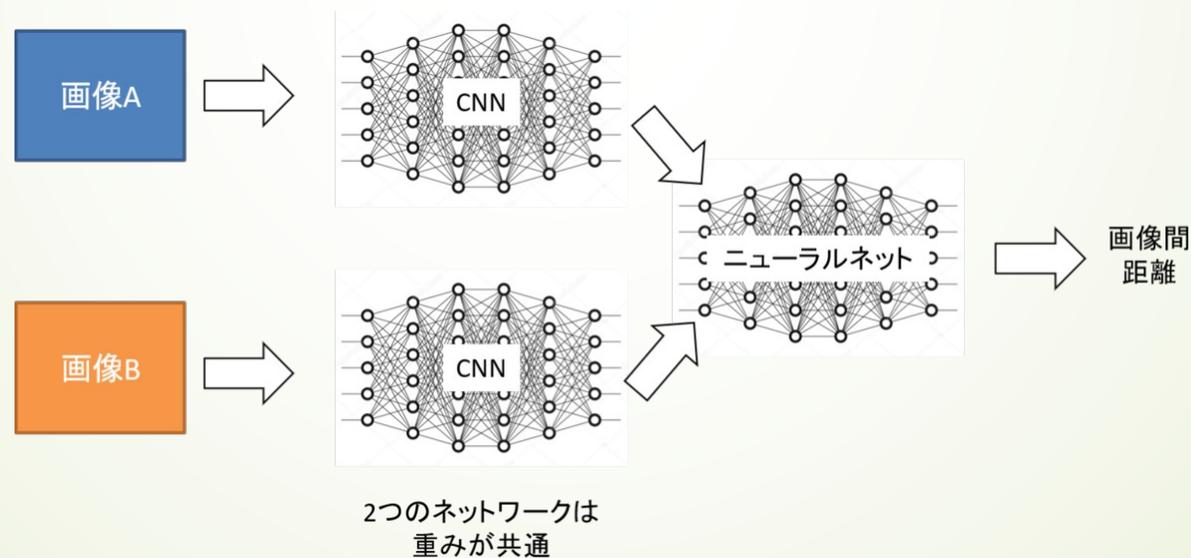
- SqueezeNet は、カリフォルニア大学バークレー校、およびスタンフォード大学の研究者によって開発されました。深さが18層の畳み込みニューラルネットワークです。100万枚を超えるイメージで学習させた事前学習済みのネットワークを、ImageNet データベースから読み込むことができます。



本研究利用したディープラーニング

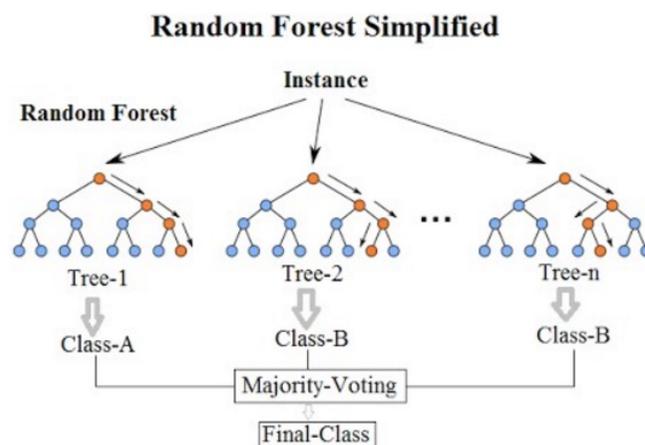
FaceNet

FaceNetはGoogleが開発された、人の顔の画像から特徴を認識するために使用されるディープニューラルネットワークです。Siamese(シヤム) Networkをベースにしており、入力画像がどのクラスに属(ぞく)するかではなく、画像同士が似ているか似てないかを表す距離を学習します。



RANDOMFORST

- ▶ ランダムフォレストは、2001年にレオ・ブレイマンによって提案された機械学習のアルゴリズムであり、分類、回帰、クラスタリングに用いられる。決定木を弱学習器とするアンサンブル学習アルゴリズムであり、この名称は、ランダムサンプリングされたトレーニングデータによって学習した多数の決定木を使用することによる。

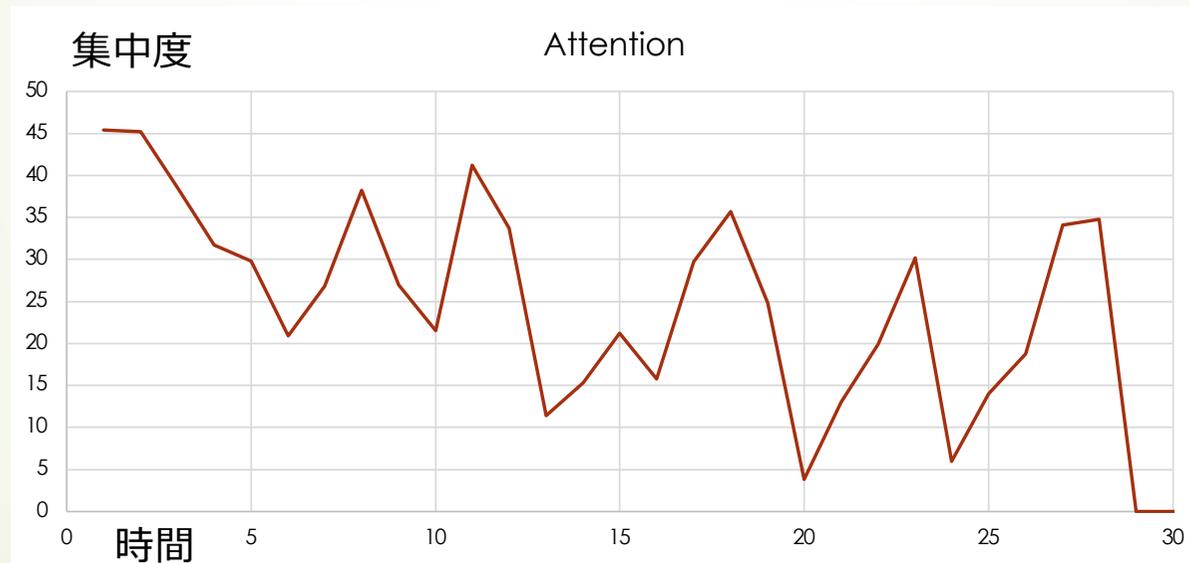


実験Aの方法

- ▶ 目的：脳波データと座圧データの関連性について
- ▶ 本研究では1人の被験者（25歳，男子留学生）に対して1回の実験を行なった。
- ▶ 被験者は脳波センサを装着し，椅子の上に座圧センサを敷（し）いた状態で，30分のゲーム（任天堂switchのmonster hunter rise）をしました。その時の座圧センサデータと脳波データを収集した。
- ▶ 座圧データに関しては5秒に1回のデータを収集するように設定し，脳波データも5秒毎にデータを保存するように設定した。

グラフの設定

グラフは脳波データによる集中度 1 分間の平均値を示している。



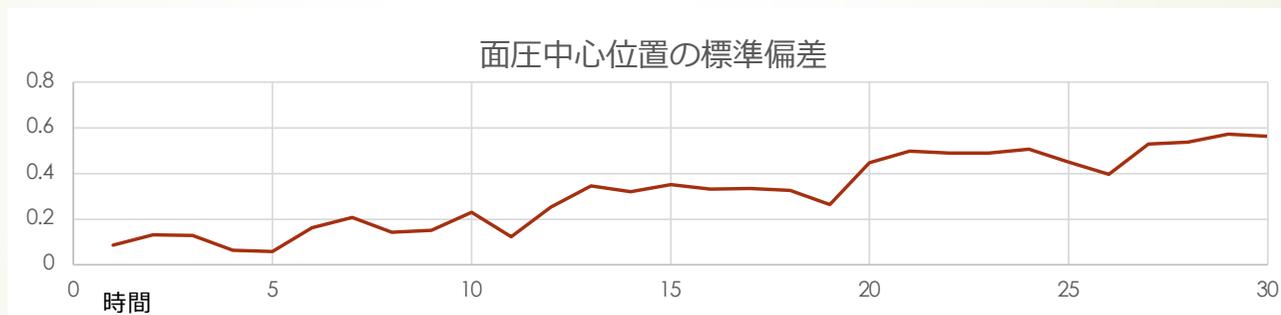
グラフの設定

グラフは座圧データによる面圧中心位置の標準偏差 1 分間の平均値を示している。



(グラフの分析)

- 時間が経過するにしたがって、座圧センサの疲労度が上がると、脳波データの集中度が下がる状態と確認された。

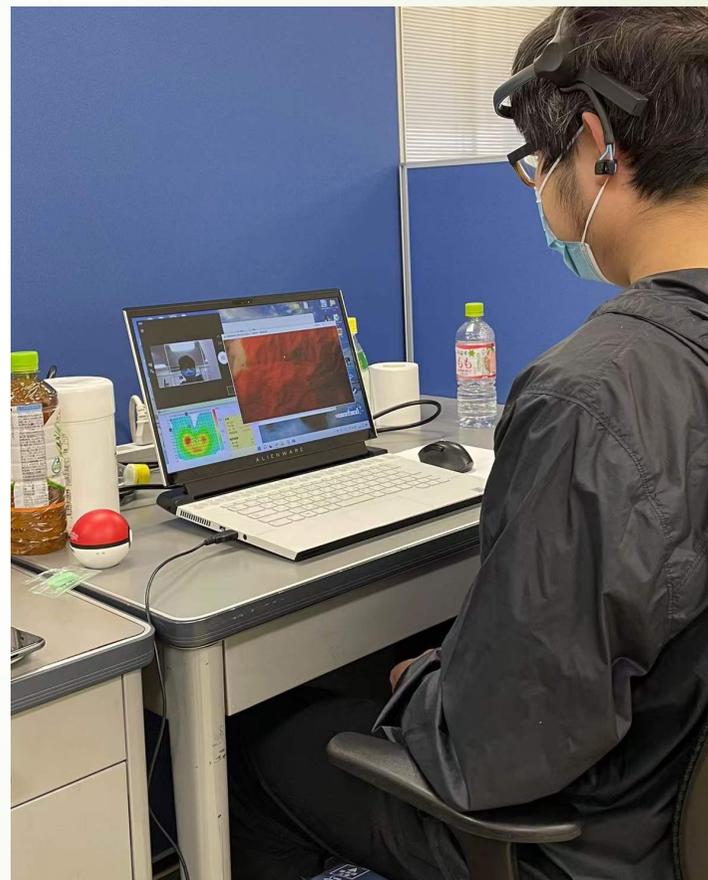
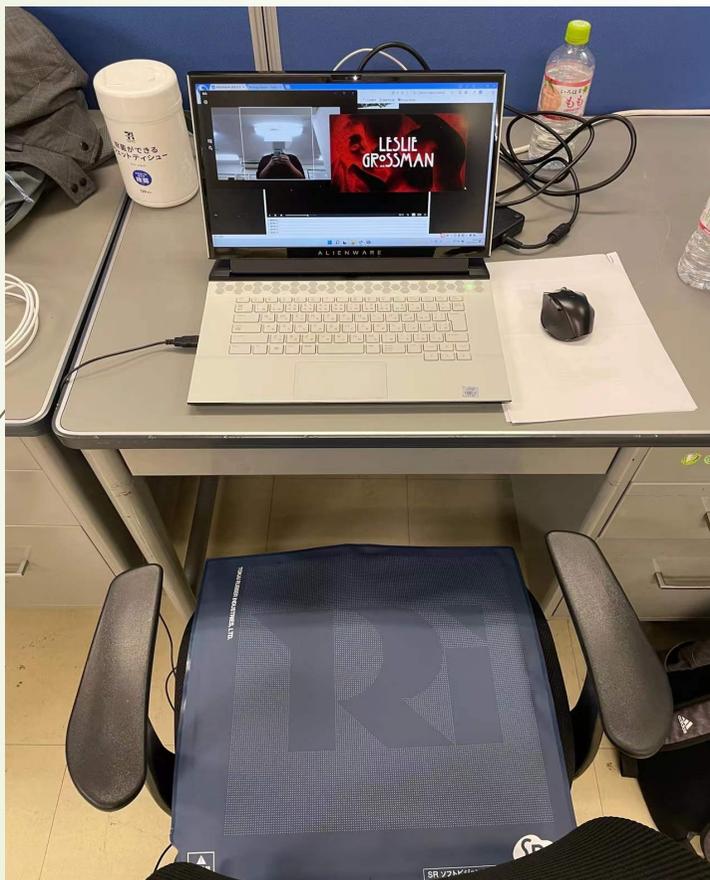


実験Bの方法

- ▶ 目的：顔画像と脳波センサ、座圧センサの関連性について。
- ▶ 本研究では1人の被験者（26歳，男子留学生）に対して2回の実験を行なった。
- ▶ 被験者は脳波センサを装着し，椅子の上に座圧センサを敷いた状態で，30分のドラマ（アメリカン・ホラー・ストーリー）を視聴してもらい，その時のセンサデータと顔画像を収集した。
- ▶ 座圧データに関しては5秒に1回のデータを収集するように設定し，顔画像も5秒毎にデータを保存するように設定した。

データ収集の様子

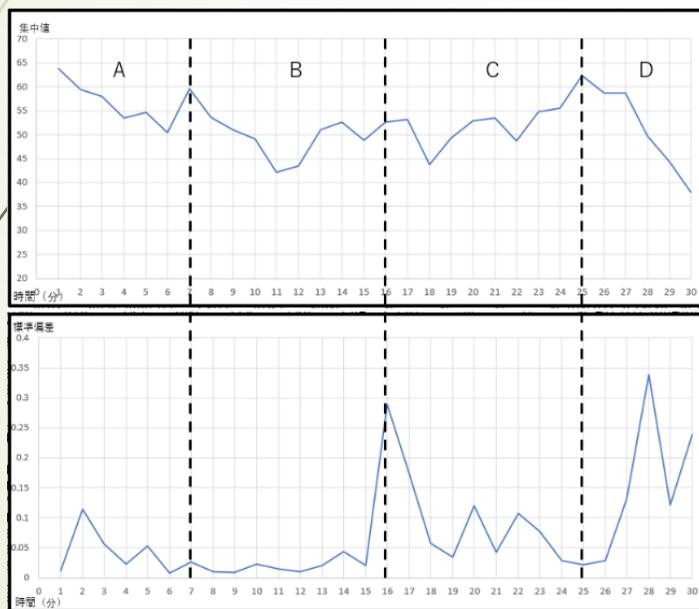
27



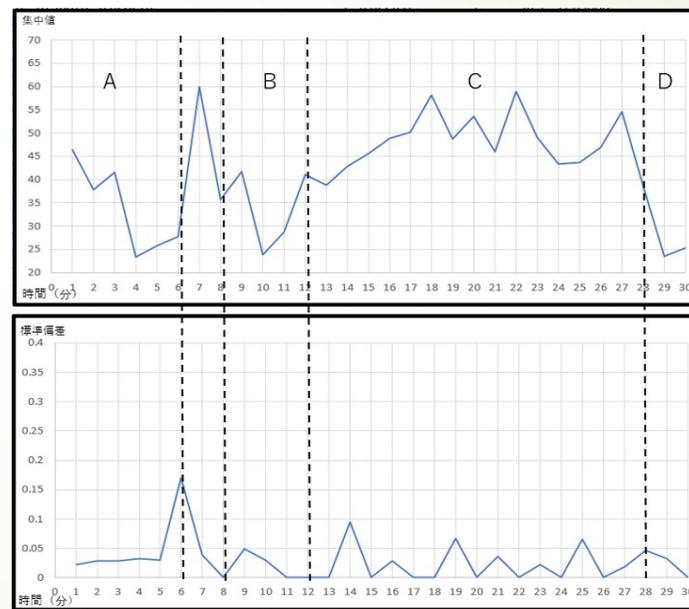
グラフの設定

28

左は1回目の実験の結果であり，右は2回目の実験の結果を示している．それぞれ上のグラフは脳波データによる集中度，下のグラフは座圧データの1分毎の標準偏差の値

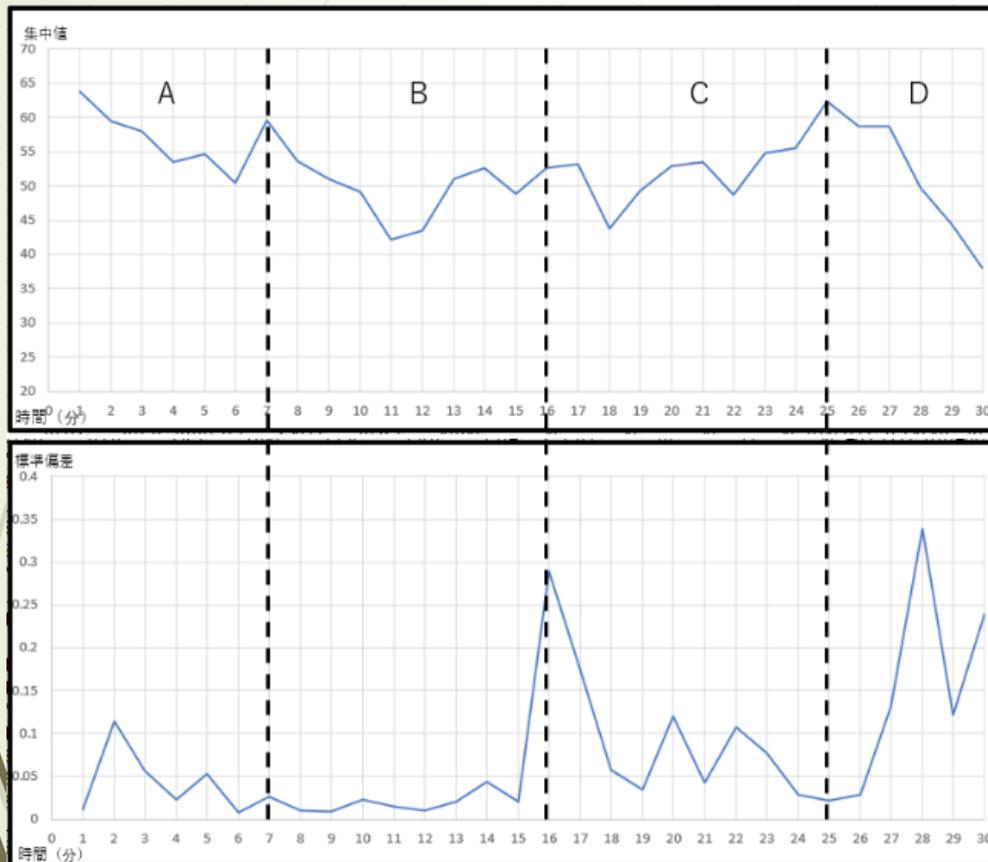


1回目の脳波データ(上)と座圧データ(下)



2回目の脳波データ(上)と座圧データ(下)

グラフの分析

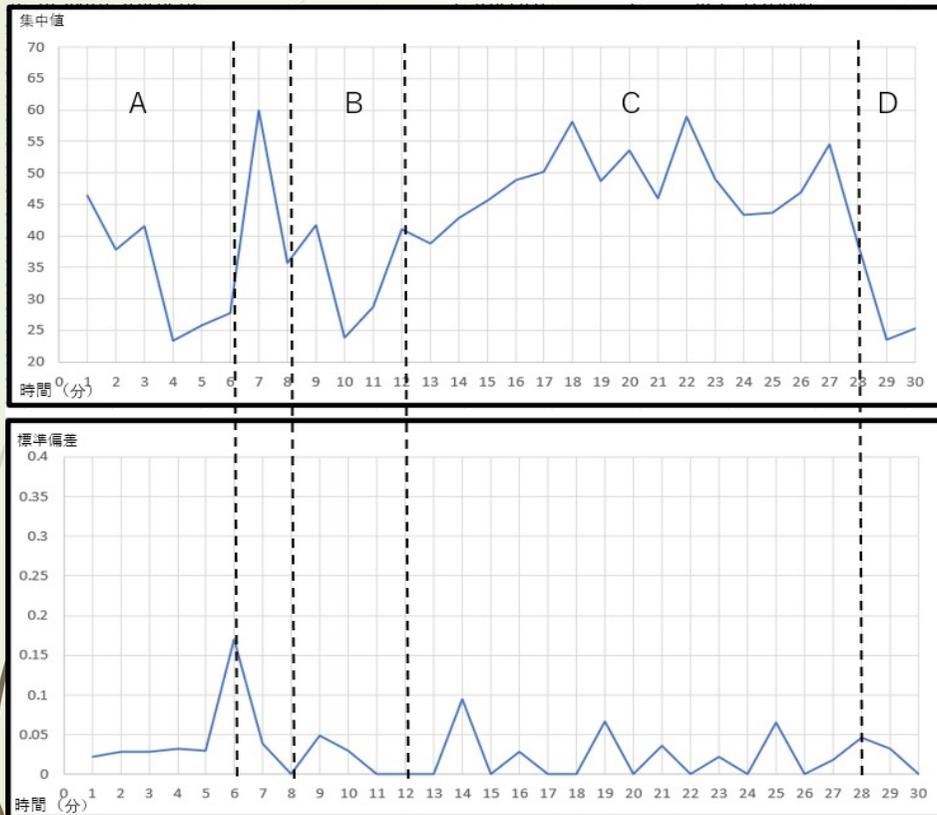


1回目の脳波データ(上)と座圧データ(下)

1回目の実験の結果を見ると、時間が経過するにしたがって、脳波による集中度は下がる傾向があり、それに伴って座圧データは大きな値が見られ、疲れる傾向が見られる。

グラフの分析

30



2回目の脳波データ(上)と座圧データ(下)

一方、2回目の結果では、1回目とは逆に、徐々に集中度の値が高くなる傾向が見られ、座圧データは、さほど変化は見られない。つまり、1回目は、最初に集中し、徐々に疲れて集中が減少するパターンであり、2回目は、最初はあまり集中できず、徐々に集中していくパターンである。

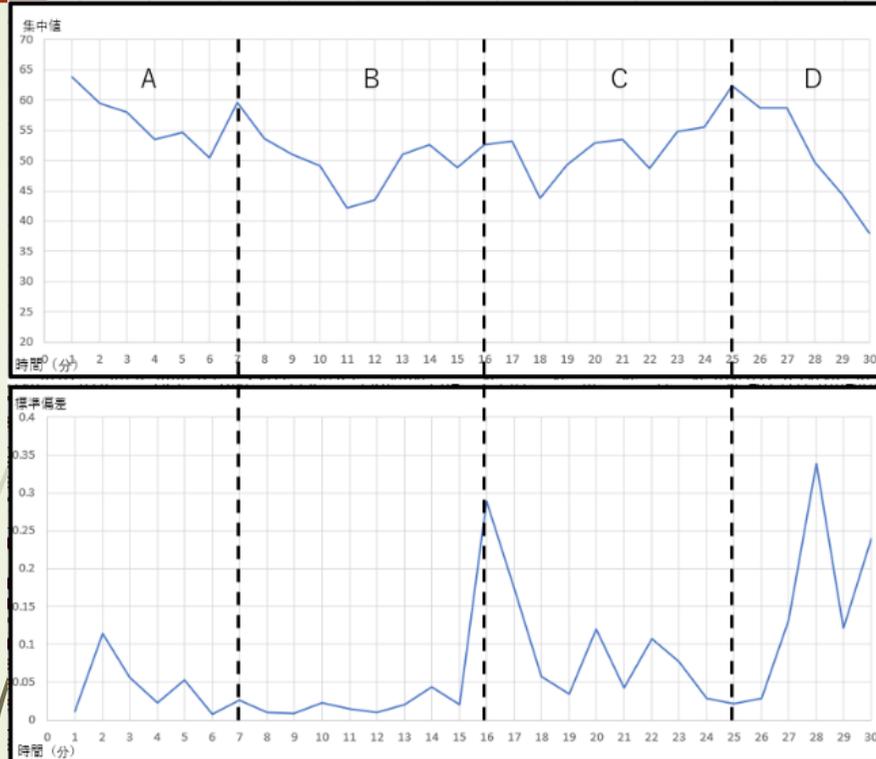
ディープラーニングによる分析

31

- 本研究では機械学習ソフトウェアのWeka を利用して、1 回目の実験の顔画像を利用して分類モデルの生成を行ない、2 回目のデータを利用して、その分類モデルの評価を実施した。（number of epochs 10;batchsize 100）

入力データの説明

32



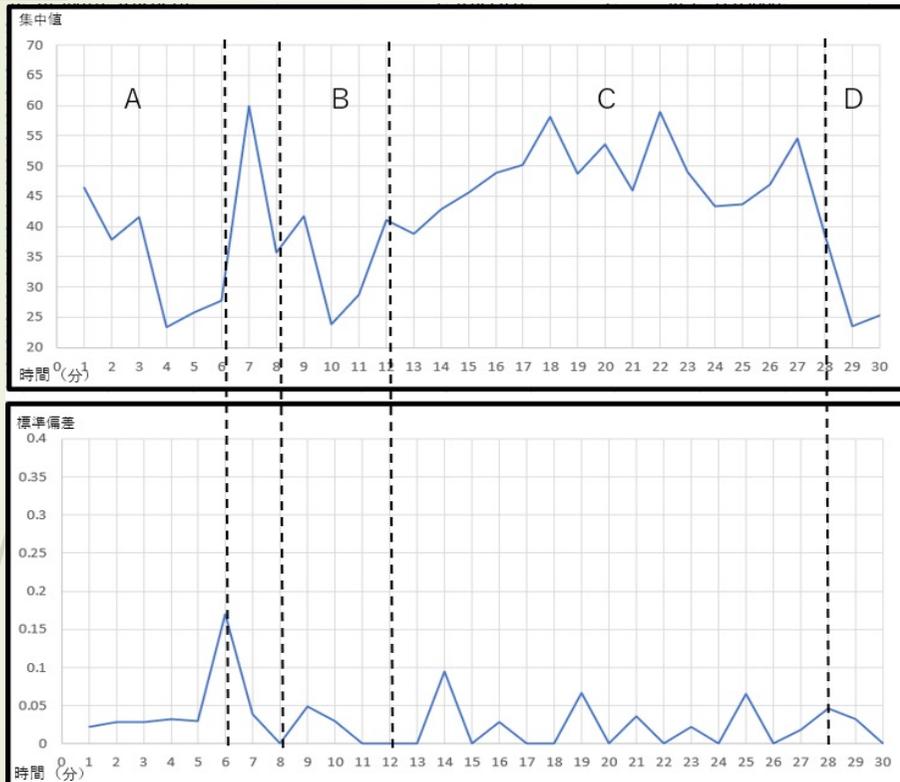
1回目の脳波データ(上)と座圧データ(下)

脳波データと座圧データを元に、収集した顔画像を、ディープラーニングによる分析を行うために、集中状態と散漫状態に分類する。

A区間は疲れもなく完全に集中した状態であり、逆に、D区間は完全に疲れていて集中が途切れ状態であるため、A区間の顔画像データ(84枚)を集中状態とし、D区間の顔画像データ(60枚)を散漫状態として分類を行なった。

入力データの説明

33



2回目の脳波データ(上)と座圧データ(下)

2回目の実験では、C区間(12-28分)は集中していると同時に疲れてない状態であり、A(0-6分)、B(8-12分)、D(28-30分)区間は、集中も途切れ、疲れも見られる状態を表しているため、C区間の顔画像データ(192枚)を集中状態とし、A、B、D区間の顔画像データ(108枚)を散漫状態として分類した。

入力データの説明

34

脳波データと座圧データを元に、収集した顔画像を、ディープラーニングによる分析を行うために、集中状態と散漫状態に分類する。

左の顔画像は集中している状態、右の顔画像は散漫状態。



集中の顔画像



散漫の顔画像

ディープラーニングによる分析

35

それぞれのモデルの分類精度を示している。LeNetは非常に精度が高く、FaceNetとSqueezeNetの精度は低い結果となった。FaceNetとSqueezeNetの精度の低い原因は、識別の特徴量が多いため、データ量が少なくいために、学習の実行が十分でなかったことが考えられる。

モデル	分類精度
LeNet	99.12
FaceNet	57.68
SqueezeNet	52.62

実験Cの方法

- ▶ 目的：OpenFaceの視線と顔画像の関連性について.
- ▶ 本研究では 2人の被験者（A: 26歳男子留学生,B: 24歳男子留学生）に対して1回の実験を行なった.
- ▶ 被験者は脳波センサを装着し，椅子の上に座圧センサを敷いた状態で，40分の授業（中国の近代歴史）を視聴してもらい，その時のセンサデータと顔画像を収集した.

データ収集の設定

37

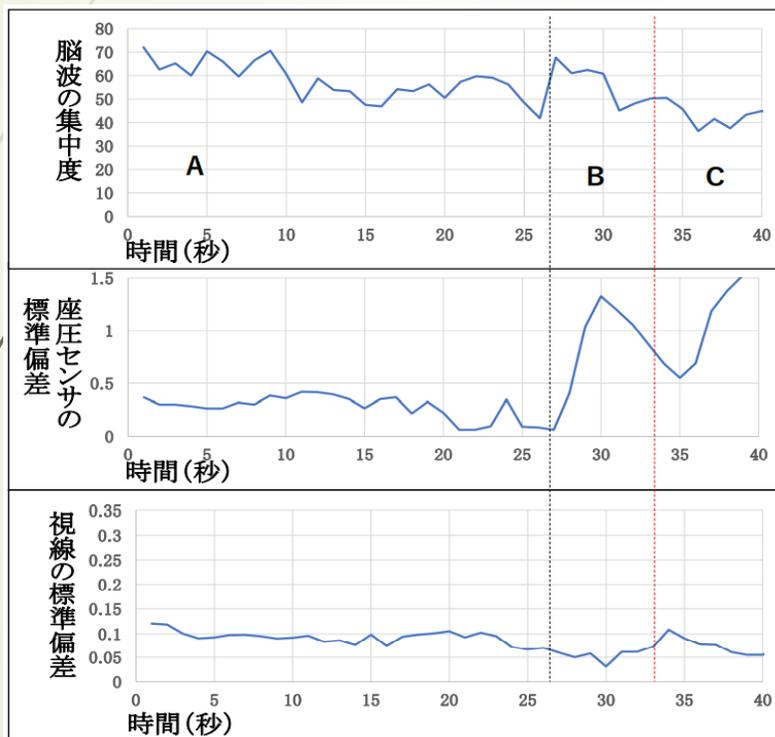
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		脳波					座圧標準偏差							視線
2	37	20.35		8.4	7.2	0.6	0.64125		0.267328	-0.06025	0.16379	-0.05079	-0.28963	0.04598
3	30	34.05		8.4	7.2	0.6	0.62375		0.286899	-0.06966	0.17828	-0.13139	-0.29377	0.14764
4	16	23.6		8.4	7.2	0.6	0.65375		0.287766	-0.07149	0.17963	-0.11042	-0.29336	0.18688
5	30	9.5		8.4	7.2	0.6	0.65		0.260632	-0.05869	0.15966	-0.08928	-0.30339	0.17782
6	13	35.225		8.4	7.2	0.6	0.65		0.258426	-0.06995	0.16419	-0.08456	-0.29403	0.14325
7	11	44.4		8.4	7.2	0.6	0.64375		0.265429	-0.04209	0.153761	-0.0848	-0.28917	0.15384
8	37	33.725		8.4	7.1	0.65	0.60875		0.243027	-0.05564	0.149332	-0.08335	-0.29661	0.15677
9	30	45.975		8.4	7.2	0.6	0.67125		0.210572	-0.10414	0.157357	0.050873	-0.29246	0.15537
10	30	54.7		8.4	7.2	0.6	0.64		0.227007	-0.09399	0.160498	0.063089	-0.17922	0.15268
11	30	47		8.4	7.1	0.65	0.67625		0.248385	-0.10649	0.17744	-0.04866	-0.07562	0.15159
12	30	48.075		8.4	7.1	0.65	0.6525		0.265095	-0.13901	0.202054	-0.06744	-0.12812	0.15109
13	23	45.55		8.4	7.2	0.6	0.65		0.258834	-0.13662	0.197727	-0.05297	-0.12258	0.14614
14	38	49		8.4	7.2	0.6	0.72		0.249143	-0.12918	0.189159	-0.00759	-0.14139	0.14713
15	48	57.55		8.4	7.2	0.6	0.6975		0.240567	-0.12461	0.182588	-0.08454	-0.16267	0.14409
16	21	42.65		8.4	7.2	0.6	0.70375		0.243033	-0.10683	0.174933	-0.13542	-0.1536	0.14642
17	44	50.425		8.4	7.2	0.6	0.705		0.245633	-0.12206	0.183845	-0.1183	-0.1441	0.12954
18	23	54.175		8.4	7.1	0.65	0.7		0.251155	-0.13492	0.193039	-0.16397	-0.14899	0.14708
19	16	40.35		8.4	7.2	0.6	0.69		0.25771	-0.13545	0.196579	-0.23882	-0.1463	0.14706
20	30	44.85		8.4	7.2	0.6	0.66875		0.261943	-0.13925	0.200596	-0.29776	-0.14819	0.15864
21	37	46.8		8.4	7.2	0.6	0.6775		0.265422	-0.15667	0.211045	-0.18668	-0.15263	0.15863
22	47	38.275		8.4	7.2	0.6	0.675		0.265704	-0.16076	0.213233	-0.1465	-0.15864	0.14899
23	38	48.5		8.4	7.1	0.65	0.73875		0.26777	-0.15421	0.210991	-0.24843	-0.14706	0.1441
24	26	40.275		8.4	7.1	0.65	0.6725		0.302106	-0.06003	0.181068	-0.21891	-0.14708	0.1536
25	26	43.225		8.4	7.1	0.65	0.69		0.277197	-0.12961	0.203402	-0.20077	-0.12954	0.16267
26	13	38.925		8.4	7.1	0.65	0.7025		0.279818	-0.10964	0.194728	-0.20569	-0.14642	0.14139
27	11	40.525		8.4	7.1	0.65	0.7		0.238392	-0.08021	0.159299	-0.1984	-0.14409	0.12258

- 座圧データに関しては1秒に1回のデータを収集し、そして1分間の面圧中心位置標準偏差を計算した
- 脳波データは1秒に1回のデータを収集し、毎分間の平均値を計算した。
- 顔画像も1秒毎にデータを保存するように設定した

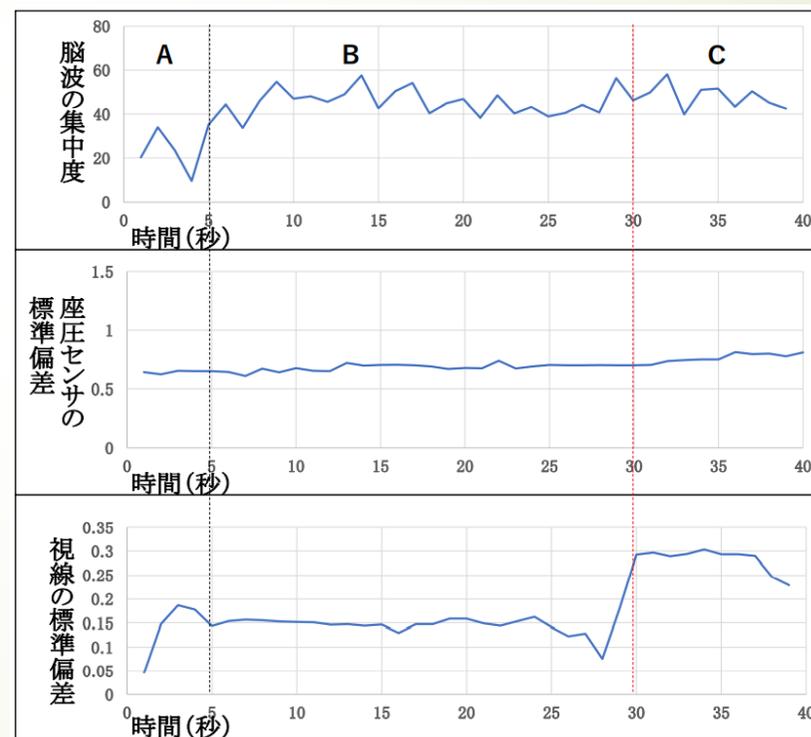
グラフの設定

38

左は被験者Aの実験の結果であり，右は被験者Bの実験の結果を示している．それぞれ上段のグラフは脳波データによる集中度，中段のグラフは座圧データの1分毎の標準偏差の値を，下段のグラフは顔画像の視線データ（視線の座標）の1分毎の標準偏差の値であ



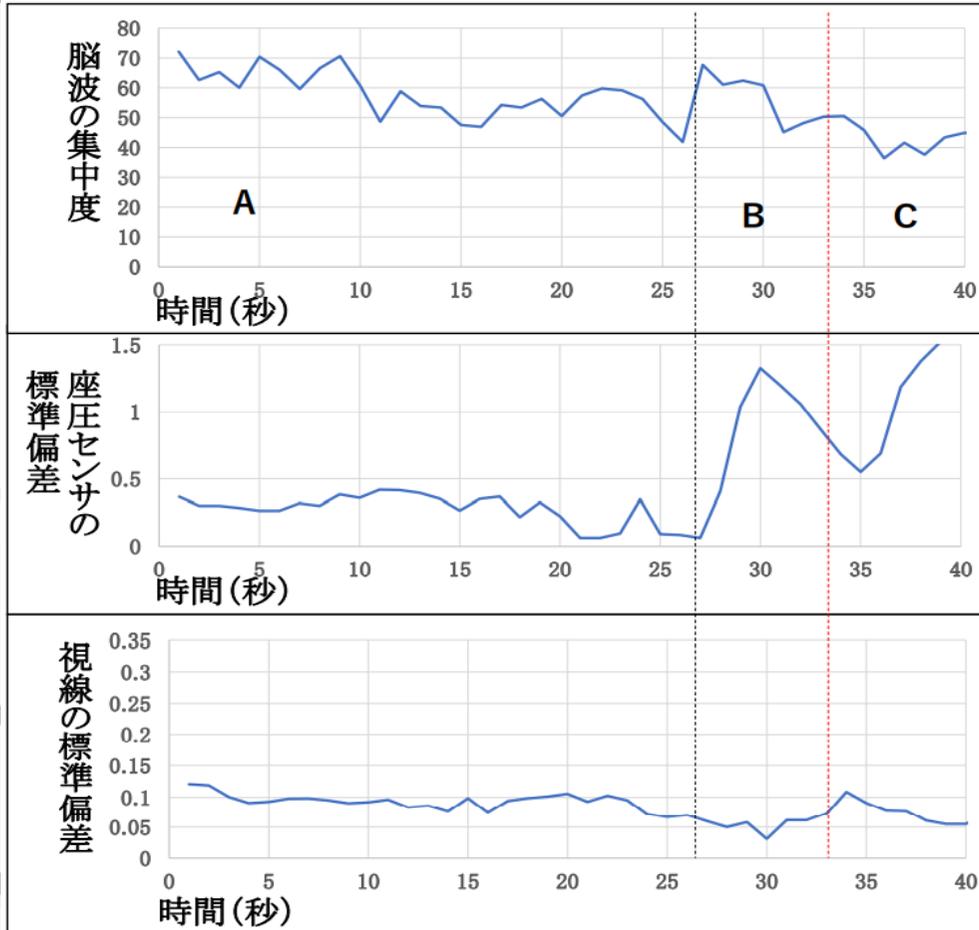
被験者Aの実験の結果のグラフ



被験者Bの実験の結果のグラフ

グラフの分析

39

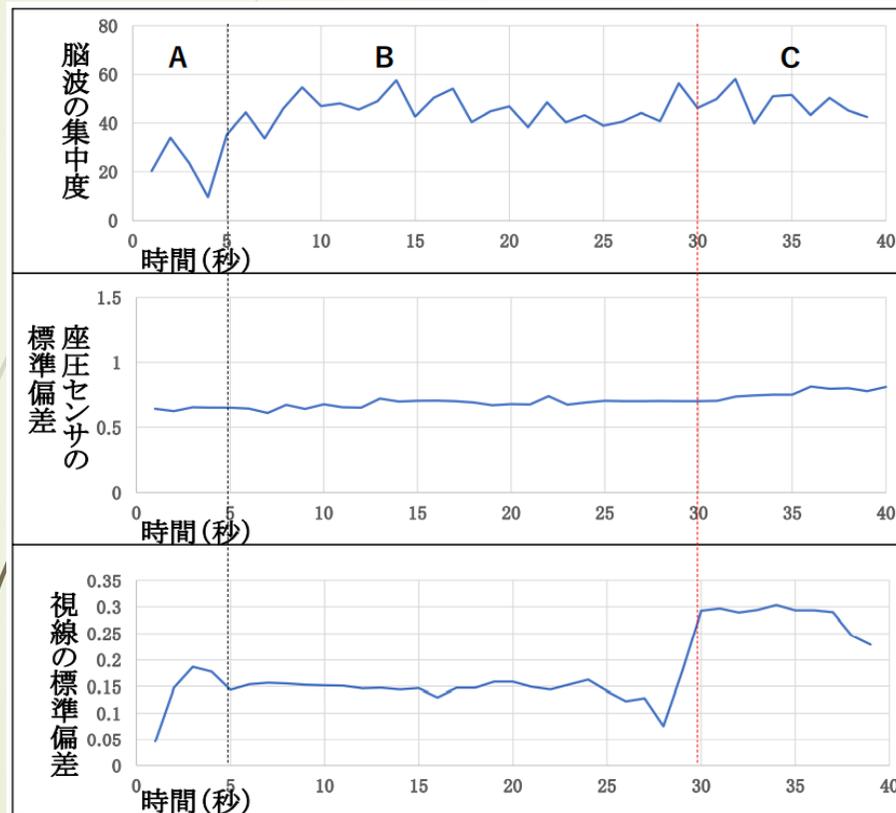


被験者Aの実験の結果のグラフ

被験者Aの実験の結果を見ると、時間が経過するにしたがって、脳波による集中度と視線の標準偏差は下がる傾向があり、それに伴って座圧データは大きな値が見られ、疲れる傾向が見られる。

グラフの分析

40



被験者Bの実験の結果のグラフ

一方、被験者Bの結果では、被験者Aとは逆に、徐々に集中度と視線の値が高くなる傾向が見られ、座圧データは、さほど変化は見られない。つまり、被験者Aは、最初に集中し、徐々に疲れて集中が減少するパターンであり、被験者Bは、最初はあまり集中できず、徐々に集中していくパターンである。

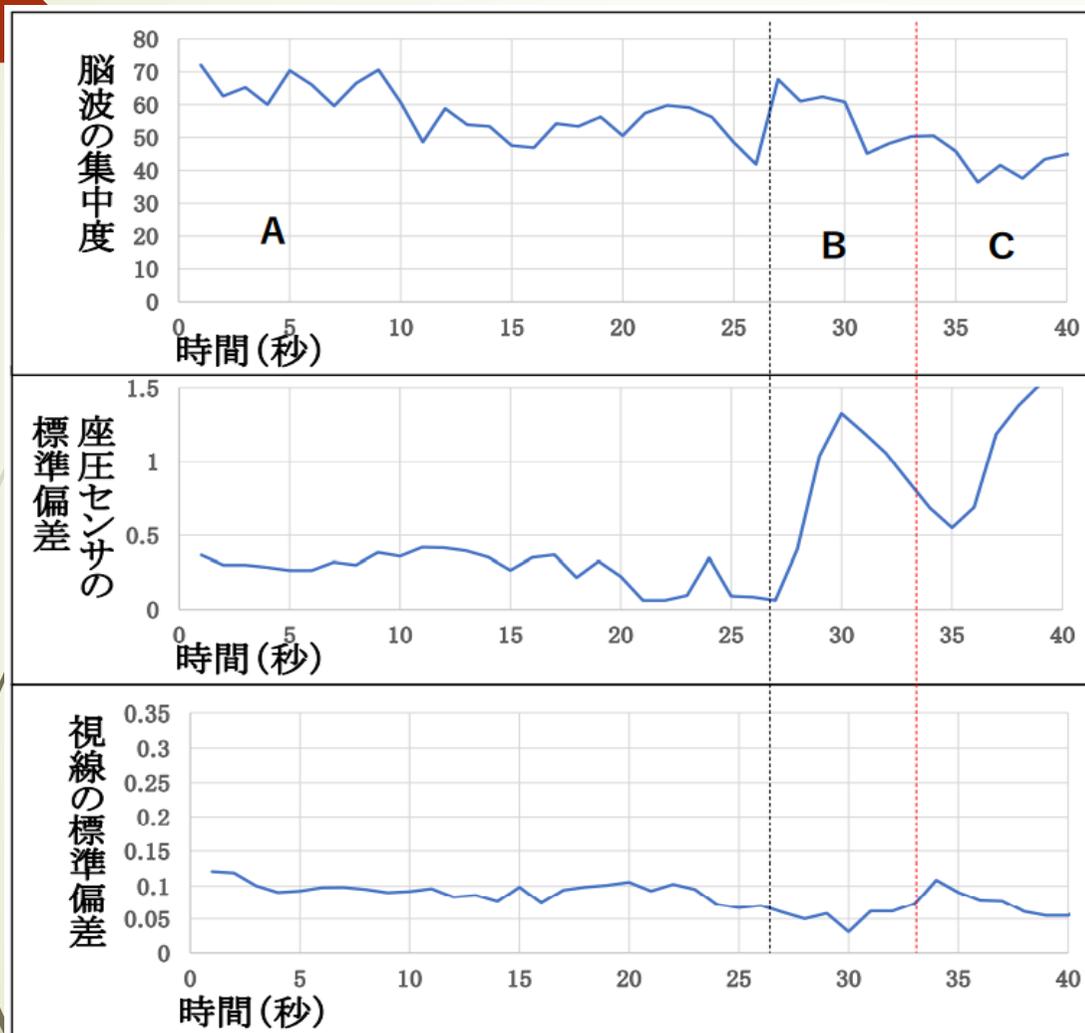
ディープラーニングによる分析

41

- 本研究では、ディープラーニングを利用し、10-fold cross validationで評価を行った。これは訓練データセットを10個のサブセットに分割し、そのうち9個で学習し残り1個で検証する方法である。そのサイクルを10回繰り返すものである。
- 本研究では、実行可能であった2つのモデルを利用して評価を実施した。

入力データの説明

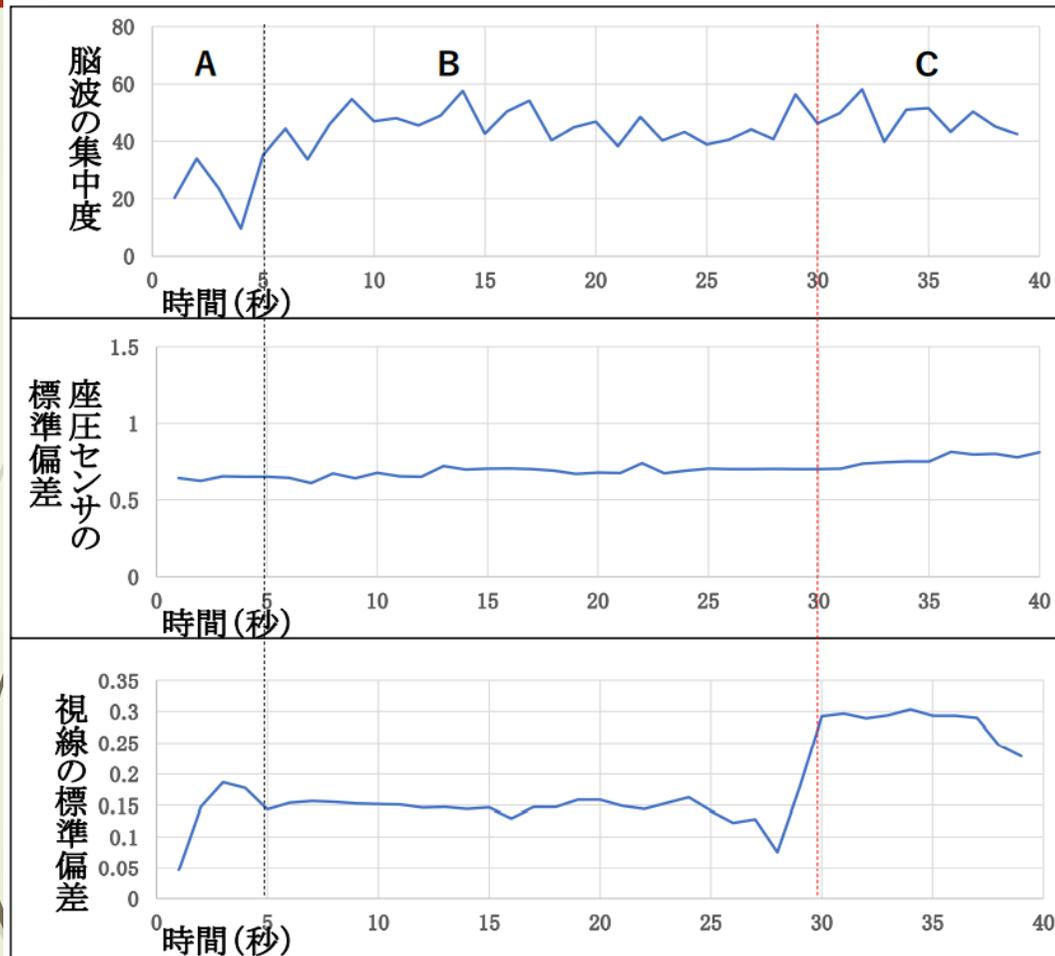
42



被験者Aにおいて、A区間（0-27分）は集中していると同時に疲れてない状態であり、B区間（27-34分）はやや疲れが見られる状態である。C区間（34-40分）は、集中も途切れ、疲れた状態を表している。したがって、A区間は疲れもなく完全に集中した状態であり、逆に、C区間は完全に疲れていて集中が途切れた状態であるため、A区間の顔画像データ（1208枚）を集中状態とし、C区間の顔画像データ（420枚）を散漫状態として分類を行なった。

入力データの説明

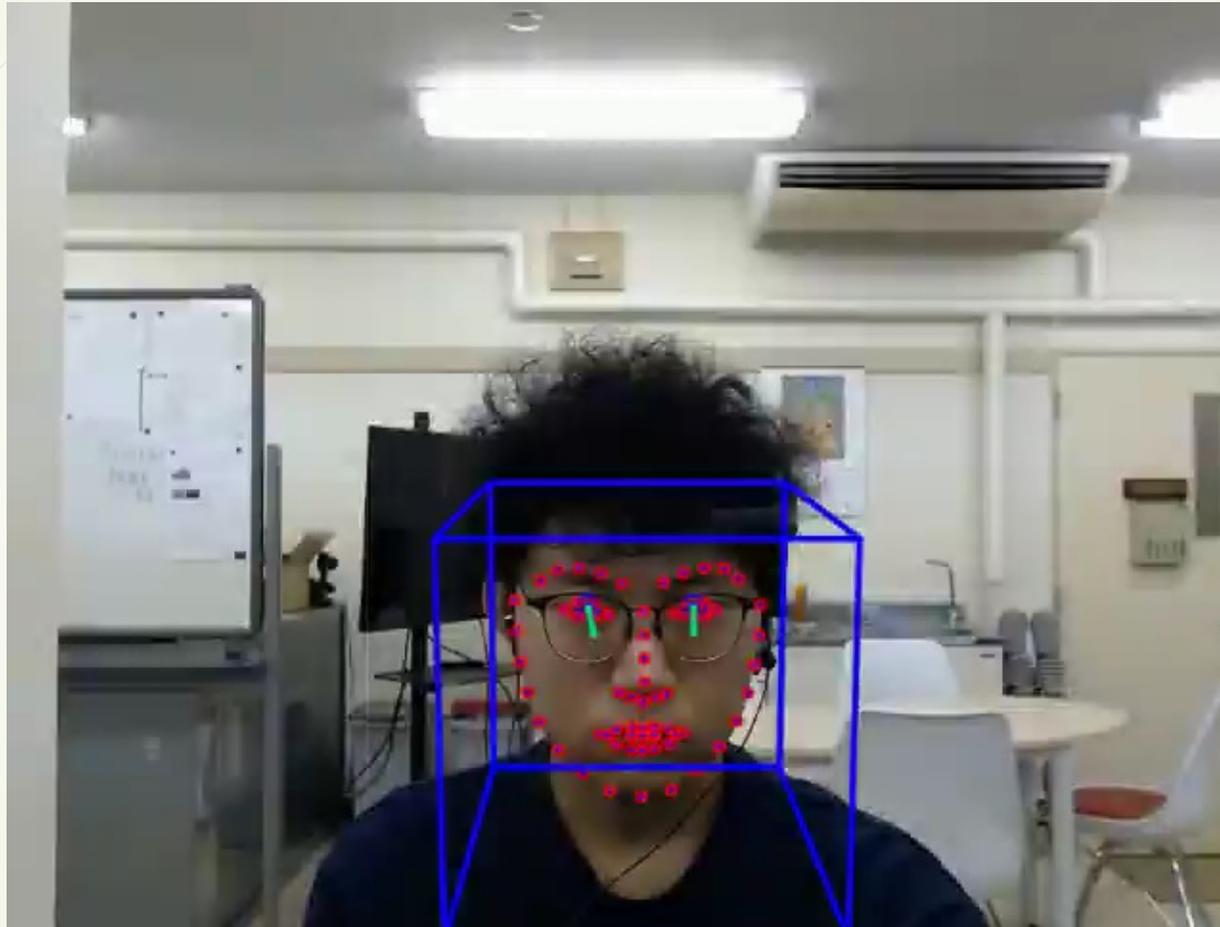
43



被験者Bでは、BとC区間（5-40分）は集中していると同時に疲れてない状態であり、A（0-5分）区間は、集中できず、疲れてない状態を表しているため、BとC区間の顔画像データ（2100枚）を集中状態とし、A区間の顔画像データ（300枚）を散漫状態として分類した。

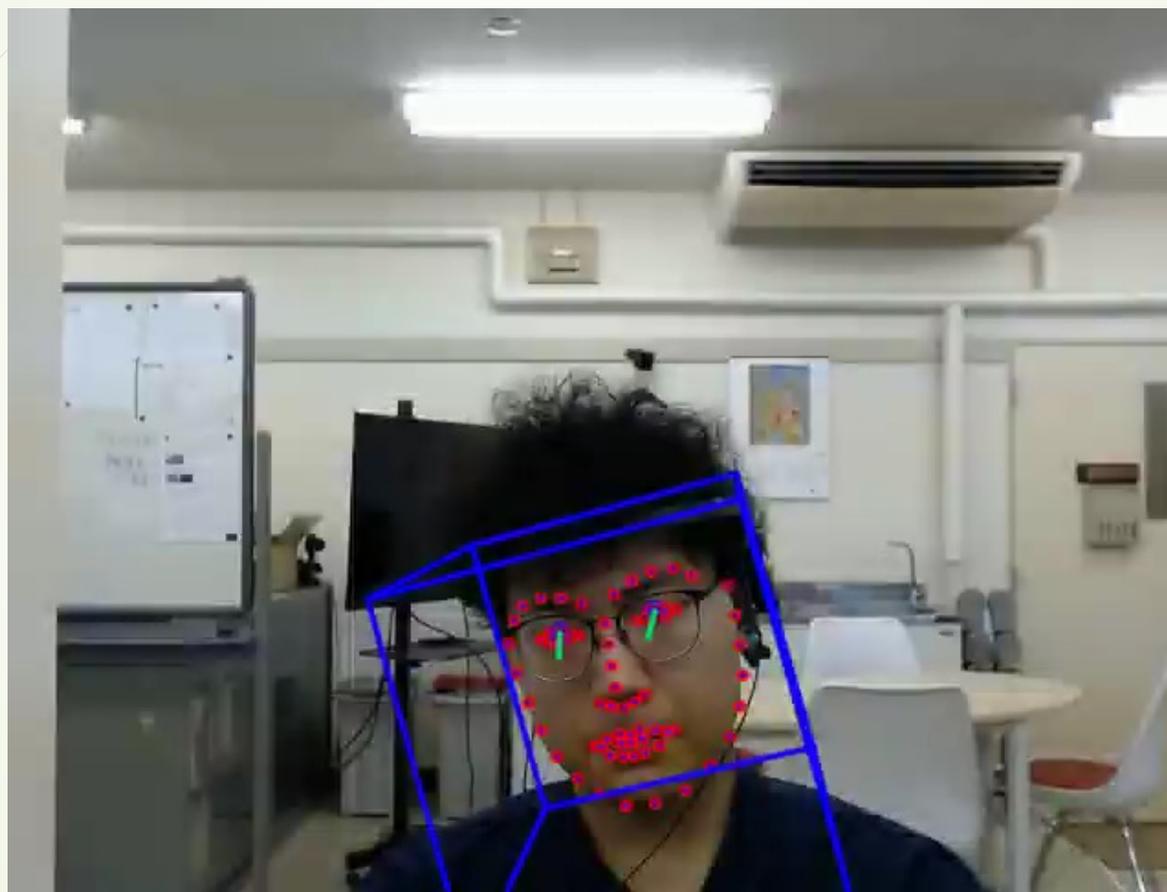
入力データの説明

44



集中している同時に疲れてないの状態の動画.

入力データの説明



集中していない同時に疲れたの状態の動画。

ディープラーニングによる分析

46

LeNet での分類精度は82.427%であった。しかし、下のグラフからほとんどの画像が集中として判断されており、散漫を散漫と判断できた割合が23/727で、3.16%とかなり低い結果となった。

下のグラフはLeNetの混合行列

	集中	散漫
集中	3299 (TP)	2 (FN)
散漫	704 (FP)	23 (TN)

視線データによる分析

機械学習の手法を利用して、被験者の集中の状態をランダムフォレストの10-fold cross validationを分析した。精度は94.87%であった。散漫の状態の分類も72.7%と比較的高い精度での分類が可能であった。

	集中	散漫
集中	90585 (TP)	2182 (FN)
散漫	3166 (FP)	8441 (TN)

まとめ

- ▶ 実験Aでは、脳波データと座圧データの関連性について分析した。
- ▶ 実験Bでは、脳波センサと座圧センサの値を基準として、集中度解析を行なった。
- ▶ 実験Cでは、脳波センサと座圧センサの値を基準として、顔画像と視線の分析を行なった。
- ▶ 顔画像と視線を分析して、集中度の判断を高い精度での分類が可能であった。